Human-level play in the game of Diplomacy by combining language models with strategic reasoning

Abstract

Despite much progress in training artificial intelligence (AI) systems to imitate human language, building agents that use language to communicate intentionally with humans in interactive environments remains a major challenge. We introduce Cicero, the first AI agent to achieve human-level performance in Diplomacy, a strategy game involving both cooperation and competition that emphasizes natural language negotiation and tactical coordination between seven players. Cicero integrates a language model with planning and reinforcement learning algorithms by inferring players' beliefs and intentions from its conversations and generating dialogue in pursuit of its plans. Across 40 games of an anonymous online Diplomacy league, Cicero achieved more than double the average score of the human players and ranked in the top 10% of participants who played more than one game. Description AI masters Diplomacy The game Diplomacy has been a major challenge for artificial intelligence (AI). Unlike other competitive games that AI has recently mastered, such as chess, Go, and poker, Diplomacy cannot be solved purely through self-play; it requires the development of an agent to understand other players' motivations and perspectives and to use natural language to negotiate complex shared plans. The Meta Fundamental AI Research Diplomacy Team (FAIR) et al. developed an agent that is able to play the full natural language form of the game and demonstrates performance well above the human average in an online Diplomacy league. The present work has far-reaching implications for the development of cooperative AI and language models for communication with people, even when interactions involve a mixture of aligned and competing interests. —YS Artificial intelligence demonstrates human-level performance in the strategic board game Diplomacy.